6:S467-76. In particular, the understanding in the art is that polyadenylation signal sequences should be reduced to enable enhanced expression of synthetic genes in plants. U.S. Pat. No. 7,741,118.

BRIEF SUMMARY OF THE DISCLOSURE

[0009] Contrary to the understanding in the art (See, e.g., U.S. Pat. No. 7,741,118), we have recently found that a reduction in the number of the polyadenylation signal sequences (e.g., AATAAA, AATAAT, AACCAA, ATATAA, AATCAA, ATACTA, ATAAAA, ATGAAA, AAGCAT, ATTAAT, ATACAT, AAAATA, ATTAAA, AATTAA, AATACA, and CATAAA) is neither necessary nor sufficient to enable enhanced expression of synthetic genes in plants. Embodiments herein make practical use of the surprising and unexpected result that preservation of these polyadenylation sequences in a synthetic coding sequence with respect to their occurrence in a native sequence may be utilized in a gene optimization process to increase heterologous protein expression.

[0010] Embodiments herein include synthetic nucleic acids encoding at least one polypeptide of interest. In embodiments, a synthetic nucleic acid encoding at least one polypeptide of interest is designed according to the constraints of specific gene design parameters that generally increase the expression of the polypeptide of interest from the nucleic acid in a host (e.g., a plant cell, plant tissue, and plant). Synthetic nucleic acid sequences may be designed from a reference nucleic acid sequence, for example, to optimize heterologous expression of the nucleic acid sequence in the host organism.

[0011] In some embodiments, a synthetic nucleic acid encoding a polypeptide of interest has been engineered for expression of the heterologous polypeptide in a host cell, wherein the polypeptide is produced in a non-genetically engineered cell in a species other than that of the host cell, and is encoded by a reference polynucleotide therein. In some embodiments, the synthetic nucleic acid is codon-optimized for expression in the host cell, for example, by altering the nucleotide sequence of the reference polynucleotide to have substantially all of the codons therein be preferred (e.g., most-preferred) codons in the host organism. In some embodiments, further analysis and engineering may be performed upon a codon-optimized synthetic nucleic acid, for example, to confirm the absence of undesired nucleic acid motifs (e.g., nucleic acid motifs forming undesirable secondary structure in an RNA molecule transcribed therefrom), confirm the absence of restriction enzyme recognition sites, and/or assure codon and sequence diversity.

[0012] In some embodiments, a synthetic nucleic acid encoding a polypeptide of interest comprises a coding sequence that is codon-optimized for expression in a heterologous cell, and at least one polyadenylation sequence selected from the group consisting of AATAAA, AATAAT, AACCAA, ATATAA, AATCAA, ATACTA, ATAAAA, ATGAAA, AAGCAT, ATTAAT, ATACAT, AAAATA, ATTAAA, AATTAA, AATACA, and CATAAA, wherein the at least one polyadenylation sequence is in the corresponding location of the coding sequence as in a reference polynucleotide. In particular embodiments, the synthetic nucleic acid comprises the same number of the foregoing polyadenylation sequences as occur in the reference polynucleotide. In particular embodiments, the synthetic nucleic acid comprises the same number of the foregoing polyadenylation sequences as occur in the reference polynucleotide, and the polyadenylation sequences are each in their corresponding locations of the coding sequence as in the reference polynucleotide.

[0013] Some embodiments include methods of making a synthetic nucleic acid encoding a polypeptide of interest, wherein the methods comprise providing the amino acid sequence of a polypeptide of interest that is encoded by a reference polynucleotide in a non-genetically engineered cell that comprises at least one polyadenylation sequence selected from the group consisting of AATAAA, AATAAT, AACCAA, ATATAA, AATCAA, ATACTA, ATAAAA, ATGAAA, AAGCAT, ATTAAT, ATACAT, AAAATA, ATTAAA, AATTAA, AATACA, and CATAAA, and producing a synthetic nucleic acid encoding the polypeptide of interest that contains at least one of the polyadenylation sequences is in the corresponding location of the coding sequence of the synthetic nucleic acid as in the reference polynucleotide. In particular embodiments, the synthetic nucleic acid is produced so that it comprises the same number of the foregoing polyadenylation sequences as occur in the reference polynucleotide, and the polyadenylation sequences are each in their corresponding locations of the coding sequence of the synthetic nucleic acid as in the reference polynucleotide.

[0014] Other embodiments include vectors (e.g., plant transformation vectors) comprising at least one of the foregoing synthetic nucleic acids. Particular embodiments include vectors comprising a transcription unit comprising a synthetic nucleic acid encoding a polypeptide of interest comprises a coding sequence that is codon-optimized for expression in a heterologous cell, and at least one polyadenylation sequence selected from the group consisting of AATAAA, AATAAT, AACCAA, ATATAA, AATCAA, ATACTA, ATAAAA, ATGAAA, AAGCAT, ATTAAT, ATACAT, AAAATA, ATTAAA, AATTAA, AATACA, and CATAAA, wherein the at least one polyadenylation sequence is in the corresponding location of the coding sequence as in a reference polynucleotide. In some examples, such a vector may comprise, for example and without limitation: a 5' non-translated sequence (e.g., comprising a plant promoter); a synthetic DNA sequence; and a 3' non-translated region (e.g., comprising a transcription termination signal).

[0015] Particular embodiments include methods of generating a plant, plant part, plant organ, plant seed, and/or plant cell that expresses a polypeptide of interest (e.g., a heterologous polypeptide of interest). Methods according to particular embodiments comprise: transforming a plant, plant part, plant organ, plant seed, and/or plant cell with at least one synthetic nucleic acid encoding a polypeptide of interest comprises a coding sequence that is codon-optimized for expression in a heterologous cell, and at least one polyadenylation sequence selected from the group consisting of AATAAA, AATAAT, AACCAA, ATATAA, AATCAA, ATACTA, ATAAAA, ATGAAA, AAGCAT, ATTAAT, ATACAT, AAAATA, ATTAAA, AATTAA, AATACA, and CATAAA, wherein the at least one polyadenylation sequence is in the corresponding location of the coding sequence as in a reference polynucleotide; and expressing the nucleic acid so as to produce the polypeptide of interest encoded thereby. In examples, a plant, plant part, plant organ, plant seed, and/or plant cell transformed with the foregoing synthetic nucleic acids may express the polypeptide of interest in a greater amount than a plant, plant part,